

PORT4NooJ

ATÉ ONDE SE PODE IR?

Anabela Barreiro*

Cristina Mota

LINGUATECA/FCCN

*CLUP

CONTEÚDO DA SESSÃO

❖ Onde estamos?

- ❑ O que é o **Port4NooJ**?
- ❑ Para que serve o **Port4NooJ**?

❖ Para onde caminhamos?

- ❑ Discussão sobre a melhoria e expansão do **Port4NooJ**
- ❑ Proposta de integração do **Port4NooJ** noutros recursos e de colaboração em projectos futuros que integrem o **Port4NooJ**

CONTEÚDO DA SESSÃO

❖ Onde estamos?

☐ O que é o **Port4NooJ**?

- Breve introdução ao sistema NooJ
- Descrição dos dicionários e gramáticas do **Port4NooJ**

☐ Para que serve o **Port4NooJ**?

❖ Para onde caminhamos?

☐ Discussão sobre a melhoria e expansão do **Port4NooJ**

☐ Proposta de integração do **Port4NooJ** noutros recursos e de colaboração em projectos futuros que integrem o **Port4NooJ**

O QUE É O PORT4NOOJ?

- Conjunto de recursos linguísticos públicos e gratuitos, que representam relações ontológicas e sintáctico-semânticas independentes do domínio

ontologia léxico-contextual – uma forma pode estar mapeada a conceitos diferentes

- Integram uma extensão bilingue para a tradução automática do português para o inglês
- Recursos que fazem mover ferramentas como o Corpógrafo, o ReWriter e o ParaMT
- Disponíveis em:

<http://www.nooj4nlp.net>

<http://www.linguateca.pt/Repositorio/Port4Nooj/>

BREVE INTRODUÇÃO AO NooJ

Características

- Plataforma de desenvolvimento de recursos linguísticos: dicionários e gramáticas são representados por transdutores de estados finitos
- Aplicação de processamento de linguagem natural: análise flexional, derivacional, sintáctica, reconhecimento de entidades mencionadas, etc.
- Implementação em C# para a tecnologia .NET
- Manipulação de diversos formatos: txt, XML, doc, pdf
- Processamento integrado de expressões regulares Perl
- Aplicação de gramáticas em cascata (TAS – Text Annotation Structure)

BREVE INTRODUÇÃO AO NooJ

QUADRO TEÓRICO

- ❑ **Léxico-gramática**, proposto por [M. Gross \(1975, 1981, 1996\)](#) na tradição da teoria transformacional de Harris
- ❑ A unidade mínima de análise e de descrição de significado é a **frase simples ou elementar** (o predicado ou núcleo da frase e os seus argumentos obrigatórios)

os nomes predicativos seleccionam os seus argumentos e os verbos suporte

- ❑ As **transformações** são **relações não orientadas** de equivalência entre frases (paráfrases)
- ❑ **Análise sistemática e detalhada** dos predicados
- ❑ **Formalização** da informação linguística
- ❑ Adequado para o **processamento automático de textos** (PLN)
- ❑ Aplicável à **tradução automática** – estudos no âmbito do LG para várias línguas

BREVE INTRODUÇÃO AO NooJ

Conceitos

- Unidade textual
- Unidades linguísticas
 - Palavras simples: *mesa, interrogar, Portugal*
 - Afixos: *re-, -izar, -ação*
 - Unidades lexicais multipalavra: *mesa redonda, de acordo com, Universidade do Porto*
 - Expressões fixas (contínuas ou descontínuas): *ir desta para melhor, não ter papas na língua, ter um bebé, ter ... em conta*
- Expressão regular
- Autómato e transdutor de estados finitos, grafo e gramática
- Tabela de léxico-gramática e transformação sintáctica

BREVE INTRODUÇÃO AO NooJ

Níveis de formalização

1. **Ortografia:** variantes ortográficas, acentos omitidos, vogais
2. **Dicionários:** vocabulário, termos técnicos
3. **Morfologia flexional:** conjugação, declinação
4. **Morfologia derivacional:** prefixação, nominalização
5. **Morfologia produtiva:** neologismos
6. **Sintaxe local:** concordância, ambiguidades
7. **Sintaxe estrutural:** árvores sintáticas, anotações
8. **Transformações:** paráfrases, análise semântica, tradução

BREVE INTRODUÇÃO AO NooJ

Funcionalidades

- Pesquisa de padrões morfológicos e léxico-sintáticos
- Construção de concordâncias
- Construção de corpos
- Construção de gramáticas encaixadas e recursivas
- Validação de gramáticas
- Anotação estruturada de corpos
- Visualização de árvores sintáticas
- Análise de expressões descontínuas
- ...

BREVE INTRODUÇÃO AO NooJ

Vantagens

- Ferramenta amiga do “não-informático” – aprende-se a construir gramáticas em 2 horas e não é necessário saber programação para criar bons recursos ou fazer uma boa investigação
 - Usada com sucesso para fins pedagógicos (em escolas secundárias, etc.)
- Permite investigação em vários domínios (biomedicina, história, desporto, etc.)
- Ambiente multilingue
- Ferramenta de acesso livre e, em breve, de código aberto
- É fácil compatibilizar com outras ferramentas
- Os dicionários têm um formato muito simples e são muito facilmente extensíveis
- Trata indistintamente unidades lexicais simples e multipalavra
- Permite criar novos níveis de anotação em corpos já anotados por outras ferramentas para finalidades distintas (ex: tradução, parafraseamento, prospecção de opinião, definição da estrutura argumental para análise de textos biomédicos, etc.)

DICIONÁRIOS DO PORT4NooJ

Categoria Gramatical
Paradigma Flexional
Atributos Sintáctico-Semânticos
Tradução em Inglês
Lema

mesa, N+FLX=CASA+CO+surf+EN=table

cair, V+FLX=ATRAIR+INMO+IntoType+EN=fall

holandês, A+FLX=INGLÊS+AN+lang+EN=Dutch

actualmente, ADV+FLX=FACILMENTE+TEMP+punc+pres+EN=nowadays

alguém, PRO+IMPERS+INDEF+EN=somebody

porque, RELINT+why+EN=why

e, CONJ+JOIN+EN=and

durante, PREP+TEMP+EN=during

cada, DET+IMPERS+INDEF+SG+EN=each

terceiro+NUM+ord+EN=one third

Amostra do dicionário geral,
representando todas as
categorias gramaticais,
formas variáveis e
invariáveis

a curto prazo, ADV+TEMP+EN=in the short run

a favor de, PREP+CAUS+EN=in favor of

cada um, PRO+INDEF+SG+EN=each one

de quem, INT+ThatType+EN=whose

quem quer que seja, REL+WhateverType+EN=whoever

além disso, CONJ+COOR+EN=besides

um quarto, NUM+frac+EN=one fourth

Amostra de compostos
invariáveis no
dicionário geral

HIV, N+FLX=PORTUGAL+AB+state+IMMUN+EN=HIV

doença maniaco-depressiva, N+FLX=CASA+AB+state+MH+EN=manic-depressive disorder

doença bipolar, N+FLX=CASA+AB+state+MH+EN=bipolar disorder

asma, N+FLX=CASA+AB+state+PULM+EN=asthma

Amostra do
dicionário de termos
biomédicos

adro da igreja, N+FLX=MENINO+PL+encl+EN=churchyard

cabo de vassoura, N+FLX=MENINO+CO+tool+EN=broomstick

bebida alcoólica, N+FLX=CASA+MA+liqu+EN=alcoholic drink+UNAMB

bebida alcoólica, N+FLX=CASA+MA+liqu+EN=booze+slang

cor de laranja, A+NAV+Apred+EN=orange

sul-americano, A+FLX=ALTO+AN+des+EN=South American

a curto prazo, ADV+LocTime+TEMP+EN=in the short run

fora de serviço, ADV+STAT+phr+EN=out of order

há muito tempo, ADV+LocTime+TEMP+puncpast+EN=a long time ago

isto é, CONJ+COOR+EN=i.e.

já não, CONJ+COOR+EN=no longer

mesmo assim, CONJ+SUB+EN=even so

juntamente com, PREP+ASSOC+EN=along

à direita de, PREP+Loc+AT+EN=at the right of

em conformidade com, PREP+ALOG+EN=in congruence with

Amostra do **DIC**ionário de
Termos e Unidades
lexicais **Multipalavra**

DicTUM

Amesterdão, N+PL+city+EN=Amsterdam

Estados Unidos da América, N+PL+count+EN=United States of America

África, N+PL+cont+EN=Africa

Extremo Oriente, N+PL+othprop+EN=Far East

Mediterrâneo, N+FLX=ANO+PL+water+EN=Mediterranean

Alpes Peninos, N+FLX=ALPES+PL+othprop+EN=Pennine Alps

ONU, N+AN+org+EN=UN

Amostra do
dicionário de Nomes
Próprios

DICIONÁRIOS DO PORT4NooJ

intimação, N+FLX=CANÇÃO+IN+inst+EN=summons
arrendamento, N+FLX=ANO+IN+inst+EN=lease
autorização, N+FLX=CANÇÃO+IN+inst+EN=fiat
autorização, N+FLX=CANÇÃO+IN+inst+EN=license
autorização, N+FLX=CANÇÃO+IN+inst+EN=permit
autorização, N+FLX=CANÇÃO+IN+inst+EN=warrant
cânone, N+FLX=ANO+IN+inst+EN=canon
cláusula, N+FLX=CASA+IN+inst+EN=clause
condição, N+FLX=CANÇÃO+IN+inst+EN=proviso
contrato, N+FLX=ANO+IN+inst+EN=contract
credo, N+FLX=ANO+IN+inst+EN=credo
declaração, N+FLX=CANÇÃO+IN+inst+EN=affidavit
decreto, N+FLX=ANO+IN+inst+EN=decree
directiva, N+FLX=CASA+IN+inst+EN=guideline
estatuto, N+FLX=ANO+IN+inst+EN=bylaw
estatuto, N+FLX=ANO+IN+inst+EN=statute
garantia, N+FLX=CASA+IN+inst+EN=guarantee
garantia, N+FLX=CASA+IN+inst+EN=warranty
intimação, N+FLX=CANÇÃO+IN+inst+EN=summons
lei, N+FLX=CASA+IN+inst+EN=law
licença, N+FLX=CASA+IN+inst+EN=licence
mandato, N+FLX=ANO+IN+inst+EN=mandate
moratória, N+FLX=CASA+IN+inst+EN=moratorium
norma, N+FLX=CASA+IN+inst+EN=norm
norma, N+FLX=CASA+IN+inst+EN=standard
ordem, N+FLX=MARGEM+IN+inst+EN=order
ordem, N+FLX=MARGEM+IN+inst+EN=ordinance
pacto, N+FLX=ANO+IN+inst+EN=pact
patente, N+FLX=CASA+IN+inst+EN=patent
renúncia, N+FLX=CASA+IN+inst+EN=waiver
subarrendamento, N+FLX=ANO+IN+inst+EN=sublease
testamento, N+FLX=ANO+IN+inst+EN=will
tratado, N+FLX=ANO+IN+inst+EN=treaty
veredicto, N+FLX=ANO+IN+inst+EN=verdict

Amostra de entradas
classificadas como Informação
+ Instrução / termo legal

UNIDADES LEXICAIS MULTIPALAVRA

dar a mão à palmatória, V+FLX=PHRDAR+EN=acknowledge being wrong

fazer o sangue subir à cabeça, V+FLX=PHRFAZER+EN=ficar tonto

ter o sangue nas guelras, V+FLX=PHRTER+EN=be alive

fazer vista grossa, V+FLX=PHRFAZER+EN=neglect

dar parte de fraco, V+FLX=PHRDAR+EN=give up

CONSTRUÇÕES COM VERBOS SUPORTE

adaptar, V+FLX=FALAR+Aux=1+INOP57+Subset132+EN=adapt+VSUP=fazer+DRV=NDRV00:CANÇÃO +NPrep=de

favor, N+FLX=MAR+Npred+AB+state+EN=favor+VSUP=fazer+NPrep=a+VRB=ajudar

literal, A+FLX=IGUAL+IN+symb+EN=literal+DRV=AVDRV05:RAPIDAMENTE

adoçar, V+FLX=COMEÇAR+Aux=1+OBJTRundif75+Subset604+EN=sweeten+DRV=ADRV11:VERDE+VCOP=tornar

transplantar, V+FLX=FALAR+Aux=1+RECTR26+Subset=504+BioMed+EN=transplant+SUBJ=AG+VSUP=fazer

+DRV=NDRV79:ANO+NPrep=de+DO=BP+IO=PAT+VSTYLE=sofrer+VSTYLE=realizar+VSTYLE=efectuar+VASP=iniciar

+VASP=prosseguir+VASP=concluir

médico, N+FLX=ANO+AN+des+med+EN=doctor

médico, N+FLX=ANO+AN+des+med+EN=physician

INFORMAÇÃO LEXICAL

- mapeamento da mesma forma a conceitos diferentes – com traduções diferentes

canto,N+FLX=ANO+CO+barr+EN=corner

canto,N+FLX=ANO+CO+barr+EN=quoin

canto,N+FLX=ANO+AB+prop+EN=birdcall

canto,N+FLX=ANO+IN+event+EN=chant

INFORMAÇÃO SINTÁCTICO-SEMÂNTICA

- Taxonomia hierárquica
- baseada na ontologia SAL da Logos SAL
- Integrada no dicionário
- Inclui informação sintáctico-semântica
- Abrange mais de 1000 categorias

ONTOLOGIA SINTÁCTICO-SEMÂNTICA

Sets and **Subsets** of the **CONCRETE** Noun Superset
Click on **CONCRETE** Superset, **sets** and **subsets** for explanations

Noun Supersets

concrete
mass
animate
place
information
abstract
process (intr)
process (tr)
measure
time
aspective

functionals

- receptacles
- bearing surfaces
- links/bridges
- thresholds, focal points, barriers
- conduits
- fasteners
- devices, tools
- cloth thing
- structural elements
- concretizations of verbals
- concretizations of mass nouns
- undifferentiated functionals
- product/brand names

* * *

agentives

- software
- vehicles
- meters
- machines/systems
- communication agents
- concrete chemical agents
- undifferentiated agentives

* * *

natural things

- minute flora
- plants
- trees
- trees/wood
- miscellaneous natural things

* * *

other concrete sets*

- impulses/lights
- blemishes/marks
- edibles (non-mass)
- edibles/color
- classifiers
- amorphous
- atomistic
- undifferentiated concrete things

* * *

*With one exception, these sets have no subsets

ONTOLOGIA SINTÁCTICO-SEMÂNTICA

ME - MEASURE Noun Sets and Subsets		
Sets and Subsets	Mnemonics (= SynSem)	Examples
abstract concepts measured by unit	ME+abs	<i>humidity, length</i>
discrete measurable concepts	ME+dis	<i>sum, increment</i>
units of measure	ME+unit	See subsets
units of weight	ME+unit+wt	<i>ounce, pound</i>
units of velocity	ME+unit+vel	<i>mph, megahertz</i>
units of volume measure	ME+unit+vol	<i>gallon, liter</i>
units of temperature	ME+unit+temp	<i>degrees celsius</i>
units of energy/force	ME+unit+ener	<i>watt, horsepower</i>
measurement systems	ME+unit+sys	<i>fahrenheit, kelvin</i>
units of duration	ME+unit+dur	<i>hour, minute, year</i>
specialized units of measure	ME+unit+spec	<i>oersted, ohm, phon</i>
units of money/value	ME+unit+value	<i>dollar, euro, forint</i>
units of linear/area measure	ME+unit+lin	<i>inch, yard, mile</i>
general undifferentiated measure	ME+undif	<i>degree, gross, share</i>

Categoria de nomes
classificados como
MEASURE (medida)

ONTOLOGIA SINTÁCTICO-SEMÂNTICA

Category

agents
 software
 concrete chemical agents
 machines/systems
vehicles
 meters
 communication agents
 functionals
 devices/tools
 fasteners
 bearing surfaces
 receptacles
 conduits
 thresholds/focal points/barriers
 links/bridges
 cloth things
 structural elements
 concretizations of verbals
 concretizations of mass nouns
 product/brand names
 natural things
 minute flora
plants
 trees
 trees/wood
 misc. natural things
 edibles (non-mass)
 edibles/color
 impulses/lights
 blemishes/marks
 classifiers
 amorphous
 atomistic

Mnemonic

CO+undagt
 CO+soft
 CO+chem
 CO+mach
CO+vehic
 CO+meter
 CO+comm
 CO+undfunc
 CO+tool
 CO+fast
 CO+surf
 CO+recp
 CO+cond
 CO+barr
 CO+link
 CO+cloth
 CO+struc
 CO+verb
 CO+mass
 CO+brand
 CO+nat
 CO+flora
CO+plant
 CO+tree
 CO+trwd
 CO+mnat
 CO+ednm
 CO+edcol
 Col+ight
 CO+blem
 CO+class
 CO+amor
 CO+atom

Examples in English

See subsets
routine
catalyst, warhead
battery, camera
truck, ship
clock, gauge
radio, radar
trinket, ornament
pliers
nail, tendon
table, shelf
bottle, barrel
chute, artery
wall, door
circuit, nerve
shirt, blanket
spar, bone
threading
acid lining
Windows NT
 See subsets
algae, spore
rose, weed
apple, willow
oak, maple
pebble, iceberg
pork chop
orange, cherry
lamp, beam
scratch, freckle
element
breeze, tide
electron, atom

Examples in Portuguese

See subsets
rotina, ficheiro
ácido sulfúrico
máquina fotográfica
automóvel
manómetro
rádio
ornamento
alicate
prego
mesa
garrafa
artéria
porta
circuito
camisola
osso

Windows NT
 See subsets
alga
erva
macieira
carvalho
iceberg
costoleta
laranja
lâmpada
sarda
elemento
brisa
átomo

Categoria de nomes
 concretos
 (CONCRETE)

FLEXÃO E DERIVAÇÃO

NooJ Community Edition

File Edit Lab Project Windows Info

Verbs.flx 121 rules.

Paradigma Flexional Nominal

plural adding -s
Logos PAT 99
ANO = <E>/m+s + s/m+p

Paradigma Flexional Verbal

regular ending in -ar (1st conjugation), like falar
FALAR = <E>/INF +
<E>/INFI+1+s + es/INFI+2+s + <E>/INFI+3+s + mos/INFI+1+p + des/INFI+2+p + em/INFI+3+p +
do/PP+m+s + dos/PP+m+p + da/PP+f+s + das/PP+f+p +
ndo/G +
<B2> (o/PR+1+s + as/PR+2+s + a/PR+3+s + amos/PR+1+p + ais/PR+2+p + am/PR+3+p) +
<B2> (ei/PS+1+s + aste/PS+2+s + ou/PS+3+s + ámos/PS+1+p + astes/PS+2+p + aram/PS+3+p) +
 (va/PI+1+s + vas/PI+2+s + va/PI+3+s + ávamos/PI+1+p + áveis/PI+2+p + vam/PI+3+p) +
<B2> (ara/PMP+1+s + aras/PMP+2+s + ara/PMP+3+s + áramos/PMP+1+p + áreis/PMP+2+p + aram/PMP+3+p) +
ei/F+1+s + ás/F+2+s + á/F+3+s + emos/F+1+p + eis/F+2+p + ão/F+3+p +
ia/C+1+s + ias/C+2+s + ia/C+3+s + íamos/C+1+p + íeis/C+2+p + iam/C+3+p +
<B2> (e/PRS+1+s + es/PRS+2+s + e/PRS+3+s + emos/PRS+1+p + eis/PRS+2+p + em/PRS+3+p) +
<B2> (asse/PIS+1+s + asses/PIS+2+s + asse/PIS+3+s + ássemos/PIS+1+p + ásseis/PIS+2+p + assem/PIS+3+p) +
<B2> (ar/FS+1+s + ares/FS+2+s + ar/FS+3+s + armos/FS+1+p + ardes/FS+2+p + arem/FS+3+p) +
<B2> (a/IMP+2+s + e/IMP+2+s + emos/IMP+1+p + ai/IMP+2+p + em/IMP+3+p) ;

Paradigma Flexional Adverbial

ending in -velmente; like amigavelmente
Logos PAT 175
AMIGAVELMENTE = <E>/Reg +
<B8>bilíssimamente/Sup ;

Paradigma Flexional para Determinante

definite article, like o
Logos PAT 151
O = <E>/m+s +
a/f+s +
s/m+p +
as/f+p ;

Paradigma Flexional Pronominal

demonstrative pronoun, like este/esse/aquele
Logos PAT 155
ESSE = <E>/m+s + s/m+p +
a/f+s + as/f+p ;

Paradigma Flexional para Pronome Interrogativo

interrogative pronoun, like qual
Logos PAT 251
QUAL = <E>/m+s + is/m+p +
<E>/f+s + is/f+p ;

Paradigma Derivacional para Nominalização

3são:
decidir > decisão
compreender > compreensão
FLX=CANÇÃO
NDRV01 =
<B3>são/N+Nom;

Prons.flx 12 rules.

OtherProns.flx 4 rules.

Noms.flx 965 rules.

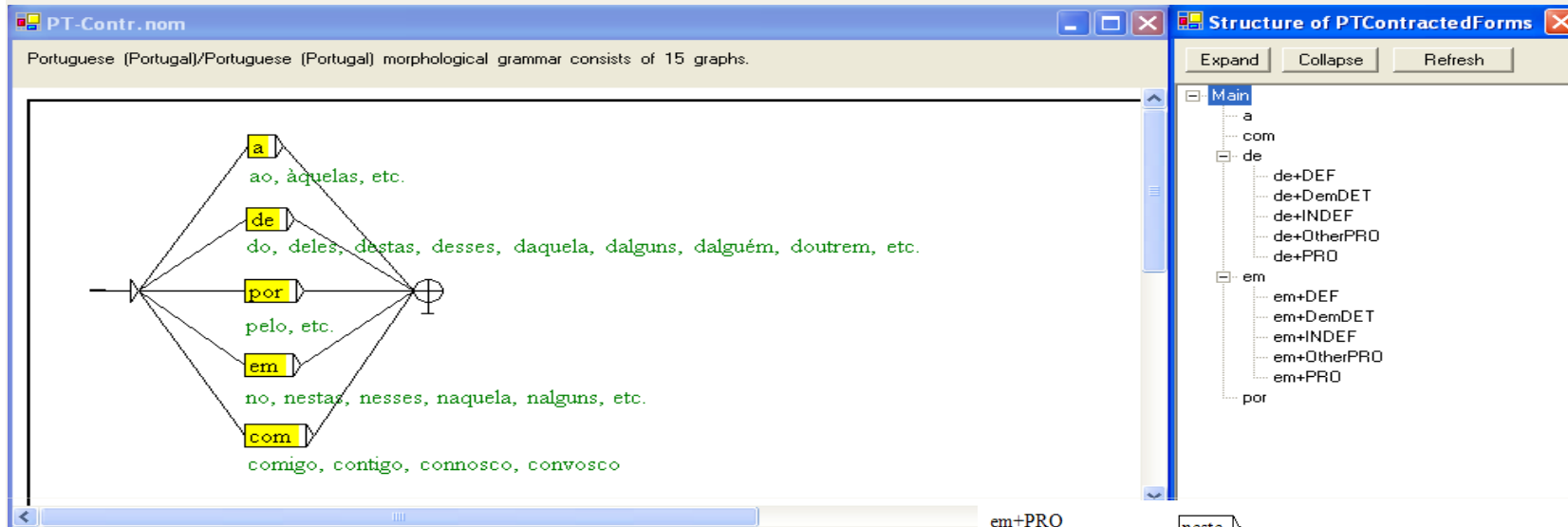
GRAMÁTICAS DO PORT4NooJ

- Gramáticas morfológicas
- Gramáticas sintáctico-semânticas
- Gramáticas de desambiguação (incluindo gramáticas sintáctico-semânticas)

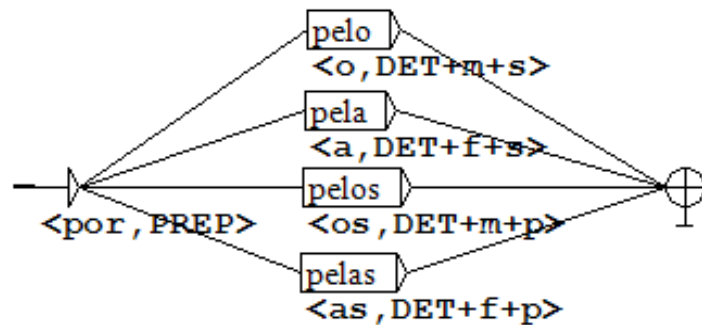
uma gramática sintáctica pode ser uma gramática de desambiguação se for usada para definir restrições sintáctico-semânticas

- Gramáticas para identificação de unidades lexicais multipalavra
- Gramáticas para parafraseamento e tradução

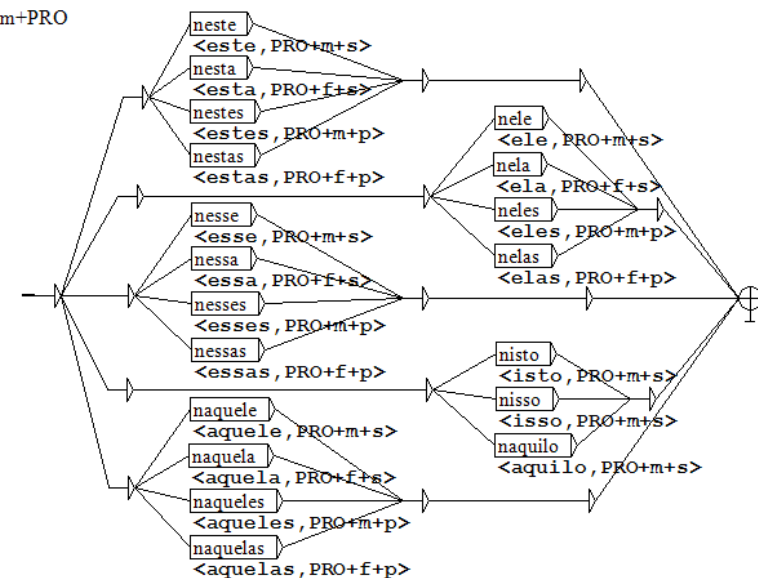
GRAMÁTICAS MORFOLÓGICAS



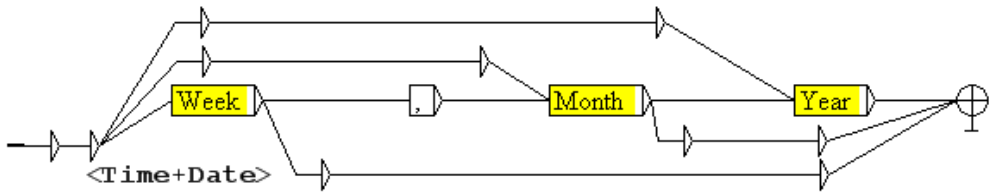
por



em+PRO

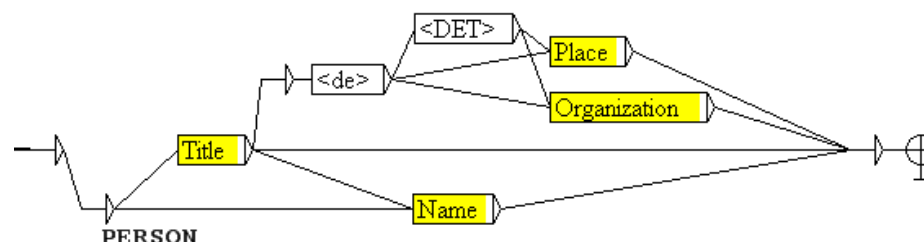


GRAMÁTICAS SINTÁCTICO-SEMÂNTICAS

São e mês de julho, graça de 1843, u... a um sol português nos primeiros dias hora da rega, por ão, da revolução de julho (isto era raiar uma alvorada	17 deste mês de julho/ <Time+Date> ano da graça de 1843/ <Time+Date> Segunda feira/ <Time+Date> de julho/ <Time+Date> de abril/ <Time+Date> meados de agosto/ <Time+Date> de julho/ <Time+Date> em 1831/ <Time+Date> de maio/ <Time+Date> no ano de 1832/ <Time+Date> Sexta feira/ <Time+Date> em 1825/ <Time+Date> sexta-feira/ <Time+Date> no último ano/ <Time+Date> ano de 1830/ <Time+Date> de agosto/ <Time+Date> sexta feira/ <Time+Date> sexta feira/ <Time+Date> sexta-feira/ <Time+Date> sexta-feira/ <Time+Date> sexta-feira/ <Time+Date> Sexta feira/ <Time+Date> sexta feira/ <Time+Date> sexta-feira/ <Time+Date> de 11 de outubro/ <Time+Date> mês de abril/ <Time+Date> de novembro/ <Time+Date> de abril/ <Time+Date> sexta-feira/ <Time+Date> sexta-feira/ <Time+Date>	, ano da graça de 1843, uma Segunda feira, dia sem nota , uma Segunda feira, dia sem nota e de boa estréia. Seis , dia sem nota e de boa estréia. Seis horas da manhã a ! A doçura que mete na alma , ondulando lascivamente com a brisa temperada , a ver-se-lhe pular os caules com a água que lhe anda p (isto era em 1831), de M. de Lafayette, de Luís Filipe, , de M. de Lafayette, de Luís Filipe, de Chateaubriand - !... Se haverá ali quem a aproveite, , uma tarde de verão como hoje calmosa, seca, mas o de aziago. , do lugar de corregedor do Ribatejo, em que já ; daí por diante em todas as sextas-feiras de cada s de Coimbra e ia formar-se em leis, quando Frei Din e Carlos que se formara no princípio daquele verão t This grammar recognizes and annotates dates  , ao por do sol, Frei Dinis aparecia no vale mais cu . - Deus tenha compaixão de , estávamos em plena e bela primavera. com o raio do sol transiente e inesperado que lhe rompe é imenso; interminável. E as últimas horas pareciam , senão não vinha eu cá, - Por qu - Não te v
--	---	--

GRAMÁTICAS SINTÁCTICO-SEMÂNTICAS

certo ponto, o	Dom Quixote/ PERSON+Title+Name
o. O que eu fiz na	Dona Branca/ PERSON+Title+Name
A	Dona Branca/ PERSON+Title+Name
esus, logo o das	Donas/ PERSON+Title
- S. Frei Gil e o	Doutor Fausto/ PERSON+Title+Name
antes e depois do	Doutor Fausto/ PERSON+Title+Name
amei já o nosso	Doutor Fausto/ PERSON+Title+Name
vê que o nosso	doutor/ PERSON+Title
impressão era a	Duquesa de Abrantes/ PERSON+Title+Place
que lhe deram. - A	Duquesa de Abrantes/ PERSON+Title+Place
adíssimo bobo de	el-rei de Dinamarca/ PERSON+Title+Place
e e o Condestável,	el-rei D. Fernando/ PERSON+Title+Name
meiro-ministro de	el-rei D. José/ PERSON+Title+Name
seu augusto amo	el-rei de Dinamarca/ PERSON+Title+Place
s promessas - como	el-rei de Prússia/ PERSON+Title+Place
o.- De como S. M.	El-Rei de Dinamarca/ PERSON+Title+Place
a, em que domina	el-rei Sancho/ PERSON+Title+Name
ssível. Desenganado	el-rei/ PERSON+Title
ho, beijou a mão à	el-rei/ PERSON+Title
nosso Carlos; e o	enfermeiro/ PERSON+Title
emada elegância do	enfermeiro/ PERSON+Title
- O hospital. - O	enfermeiro/ PERSON+Title
parte da nossa	Espanha/+Place
rasedidática das	Espanhas/+Place



de bivaque, o soldado que lhe chamou maluco ao pensad
 Mas em meia hora de conversação, de trato, descobriam-
 - Chega- se enfim ao Vale de Santarém.

, o que alguns anos depois ressuscitou em Sterne com tã
 e a Rainha D. Leonor, Camões desterrado aqui, Frei Lu
 ? Por onde está Ixião e Tântalo, por onde demora Sísifo
 Por pouco mais que se generalize o princípio,
 prometeu uma constituição; e não faltou ainda, porque,
 tinha menos juízo do que Yorick, seu bobo. - Doutrina deste

Depois há

de que um poder sobre-humano não permitia que ele s
 , e daí tomou um dia o caminho de Santarém, chegou àquel
 que o velava, uma bela mulher de estatura não acima de
 que o velava. O quarto era com efei

- Georgina.

é, geologicamente falando, já tão África, tão pouco
 chamou romance em endechas. Eu, adotando para ele,

GRAMÁTICAS DE DESAMBIGUAÇÃO

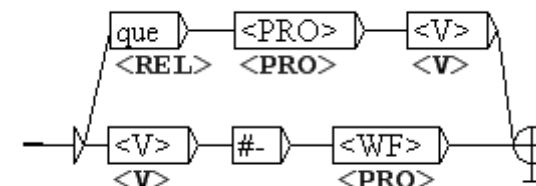
utor prestou-se a
gi-la ele mesmo,
rrigiu-a, aditou-a,
to; e como resolveu
e um bom charuto.
cias de um amigo,
por isso mesmo vou:
das populações, e
ita reação antes de
ossos charutos, e
u cigarro de papel,
ribatejano.
raros pintos por
e forçado, assim
i a nossa gente
, e as ricas terras
- A força é
estão em terreno
inha. - A força é
um homem do campo

civilização: e
r ao leitor, para
moda germânica,
em tudo e com
, e que pode bem
em personalizar-se,
utopias, pode bem
porque receio muito

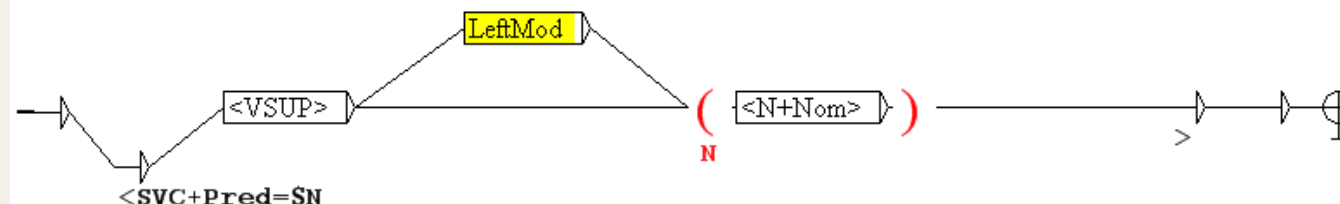
dirigi-la/<V><PRO>
corrigiu-a/<V><PRO>
aditou-a/<V><PRO>
imortalizar-se/<V><PRO>
Travam-se/<V><PRO>
decidem-me/<V><PRO>
pronunciei-me/<V><PRO>
que se chama/<REL><PRO><V>
completar-se/<V><PRO>
deixe-mos/<V><PRO>
que me vai/<REL><PRO><V>
Acenderam-se/<V><PRO>
que se manifesta/<REL><PRO><V>
que o viu/<REL><PRO><V>
que o satchou/<REL><PRO><V>
que lhes levam/<REL><PRO><V>
que se fala/<REL><PRO><V>
que lhe convinha/<REL><PRO><V>
que se fala/<REL><PRO><V>
que se deita/<REL><PRO><V>
Declaram-se/<V><PRO>
mostra-se/<V><PRO>
que ele esteja/<REL><PRO><V>
que se mete/<REL><PRO><V>
que se explica/<REL><PRO><V>
personalizar-se/<V><PRO>
simbolizar-se/<V><PRO>
representar-se/<V><PRO>
que se esqueça/<REL><PRO><V>
Parece-me/<V><PRO>

ele mesmo, corrigiu-a, aditou-a, alterou-a e
aditou-a, alterou-a em muitas partes, e a i
em muitas partes, e a ilustrou com as notas
escrevendo estas suas viagens. Parte pa
de razões os ilhaves e os Bordas- d'Águas
as tonteiras de um jornal, que por mexeriqui
São 17 deste mês de julho,
a si mesma por excelência a Sociedade, os

No entreta
os precintos aristocráticos da ré; à proa, que é país d
emprestar lume. - Dou-lho eu, senhor
os charutos, e atentamos mais devagar na companhia
o sempre clamoroso e sempre vazio entusiasmo das multid
) Isto é um fidalgo como se quer. Nunca o vi numa ferr
e plantou, e o fez o que é, e fez terra das areias da charn
as enchentes. Mas nos , pé no barco, pé na terra, tã
- tornou o campino para estabelecer a questão em t
- A força é que se fala: um homem do campo que se deita
um homem do campo que se deita ali à cernelha de um toi
ali à cernelha de um toiro que uma companhia inteira de
típicas, simbólicas e míticas estas viagens. Faz o A.
como ela é dirigida pelo cavaleiro da Mancha, D.
prevenido; não cuide que são quaisquer dessas rabis
hoje em tudo e com que se explica tudo... quanto se nã
tudo... quanto se não sabe explicar.
, simbolizar-se pelo famoso mito do cavaleiro da mancha, D.
pelo famoso mito do cavaleiro da mancha, D. Quixote;
pela rotunda e anafada presença do nosso amigo velho,
Somos chegados ao triste desembarc
estar mais deserto e sujo, mais abandonado e em ruín



GRAMÁTICAS PARA RECONHECIMENTO E ANOTAÇÃO DE UNIDADES LEXICAIS MULTIPALAVRA

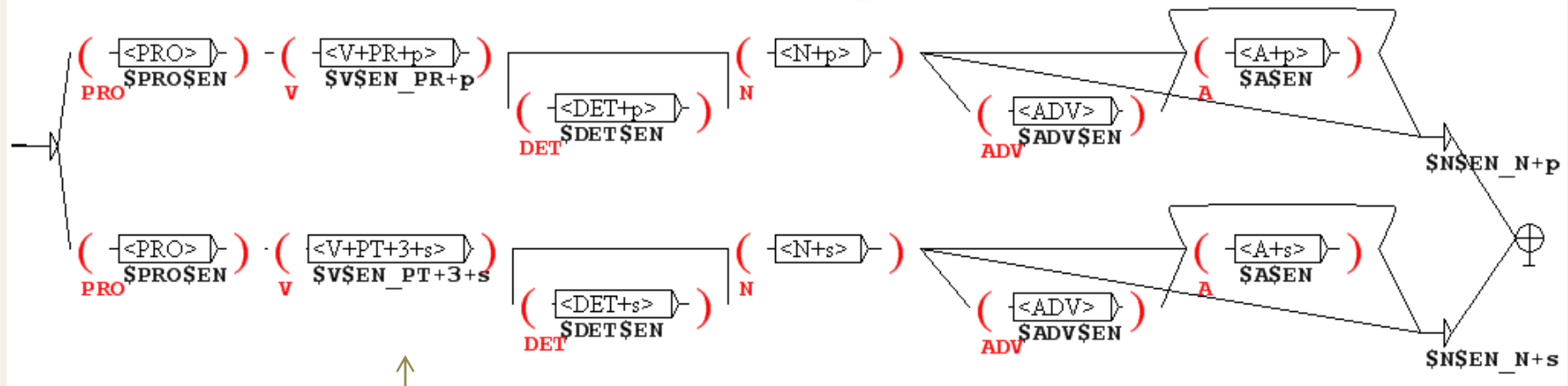


no bordado para lhe
 uer sair. -- Está a
 a uma agulha me faz
 à primeira e tem de
 o ponto de me fazer
 fazer dar um grito.
 lições a iniciados.
 regados, alunos que
 ianças de colo, que
 que dão graças por
 primeira pedra, que
 jogador encantador,
 a defender, como se
 a maneira de evitar
 er o pino, pôr-se a

dar um beijo/<SVC+Pred=beijo>
dar uma festa/<SVC+Pred=festa>
dar um salto/<SVC+Pred=salto>
fazer várias tentativas/<SVC+Pred=tentativas>
dar um grito/<SVC+Pred=grito>
Toma conta/<SVC+Pred=conta>
Toma conta/<SVC+Pred=conta>
fazem gazeta/<SVC+Pred=gazeta>
dão graças/<SVC+Pred=graças>
terem este local acolhedor/<SVC+Pred=local>
dá a outra/<SVC+Pred=a>
dá gosto/<SVC+Pred=gosto>
tivesse a bola atada/<SVC+Pred=bola>
fazer o pino/<SVC+Pred=pino>
dar murros/<SVC+Pred=murros>

na cara e os nossos olhos se cr
? -- perguntou. Talvez fosse do
; e, quando não consegue encor
-- o que é muito raro --, fica m
. Toma conta da loja quando o j
da loja quando o pai está a dar
da loja quando o pai está a dar
e mães com crianças de colo, q
por terem este local acolhedor
e alegre para passarem as tarde
face, e por fora, mas não me co
vê-lo a defender, como se tives
aos pés, desviando os adversár
, pôr-se a dar murros no peito c
no peito ou gritar de excitação.

GRAMÁTICAS PARA PARAFRASEAMENTO E TRADUÇÃO



Grafo para traduzir frases simples

Tradução e parafraseamento de frases simples

Concordance for Text PTSimpleSentences-3pl.not

Clear Concordance 4 characters before, and 5 after. Display: ☒ Inputs ☒ Outputs

Text	Before	Seq.	After
Elas visitam cidades.	Elas são homens honestos/they are honest men		. Elas são mulheres simpáticas. Elas
Elas visitam cidades.	Elas são homens honestos/they are straightforward men		. Elas são mulheres simpáticas. Elas
comida. Eles vendem sapatos.	Elas compram casas/they purchase houses		. Elas sonham acordados. Eles sonham
comida. Eles vendem sapatos.	Elas compram casas/they purchase homes		. Elas sonham acordados. Eles sonham
comida. Eles vendem sapatos.	Elas compram casas/they buy houses		. Elas sonham acordados. Eles sonham
comida. Eles vendem sapatos.	Elas compram casas/they buy homes		. Elas sonham acordados. Eles sonham

DADOS QUANTITATIVOS DO **PORT4NooJ**

Por categoria gramatical

	Simples	Compostos	Domínio biomédico	TOTAL
Verbos	8887	847		9734
Nomes	24620	12256	340	37216
Adjectivos	18469			18469
Advérbios	4753			4753
Determinantes	268			268
Preposições	288			288
Pronomes	192			192
Conjunções	168			168
Expressões numéricas	203			203
TOTAL				71291

DADOS QUANTITATIVOS DO **Port4NooJ**

Recursos parafrásticos do **Port4NooJ**

Relações semânticas	
Verbo – Nome predicativo	8472
Adjectivo predicativo - Advérbio	222
Total de relações	8694

CONTEÚDO DA SESSÃO

❖ Onde estamos?

✓ O que é o **Port4NooJ**?

❑ Para que serve o **Port4NooJ**?

- Identificação e anotação de padrões linguísticos
- Análise e transformação sintáctica
- Tradução automática

❖ Para onde caminhamos?

❑ Discussão sobre a melhoria e expansão do **Port4NooJ**

❑ Proposta de integração do **Port4NooJ** noutros recursos e de colaboração em projectos futuros que integrem o **Port4NooJ**

PARA QUE SERVE O PORT4NOOJ?

- Anotação linguística e processamento de corpos em português
- Caixa de ferramentas para testar e apoiar a investigação linguística teórica em português
- Base para aplicações de tecnologia de linguagem humana (HLT) em português, incluindo a TA do português para o inglês

FERRAMENTAS QUE UTILIZAM O PORT4NOOJ

➤ **Corpógrafo**

- Serviço na rede para a construção de corpos e extracção e tratamento de terminologia
- Disponível em: <http://www.linguateca.pt/Corpografo/>

➤ **ReWriter**

- Parafraseador para a pré-edição de textos
- Versão portuguesa **ReEscreve** – serviço público disponível em: <http://www.linguateca.pt/ReEscreve/>

➤ **ParaMT**

- Parafraseador bilingue / multilingue (tradutor) para sistemas de tradução automática

CONTEÚDO DA SESSÃO

❖ Onde estamos?

- ✓ O que é o **Port4NooJ**?
- ✓ Para que serve o **Port4NooJ**?

❖ Para onde caminhamos?

- ❑ Discussão sobre a melhoria e expansão do **Port4NooJ**
- ❑ Proposta de integração do **Port4NooJ** noutros recursos e de colaboração em projectos futuros que integrem o **Port4NooJ**

COMO É QUE O **PORT4NOOJ** PODE EVOLUIR?

- Alargamento dos recursos
 - Dicionários e gramáticas para o reconhecimento de unidade lexicais multipalavra
 - Adaptação e integração de recursos de domínio público – JSpell, entidades mencionadas, etc.
 - Inclusão de sinónimos
- Alargamento e melhoria da ontologia e relações sintáctico-semânticas
- Estandardização de anotações
- Correção de erros “hereditários”
- Melhoria das descrições flexionais dos compostos
- Alargamento das descrições derivacionais

COMO É QUE O **PORT4NOOJ** PODE EVOLUIR?

- Alargamento das relações sintáctico-semânticas entre predicados e elementos morfossintáctica e semanticamente relacionados

Exs: *apresentar – apresentação – apresentado*
 literal – literalmente

- Melhoria e alargamento das capacidades parafrásticas por detrás do ReWriter
 - Alargamento das paráfrases de construções com verbos suporte
 - Parafraseamento de vários outros fenómenos linguísticos – inclusão de variação estilística e controlo da qualidade linguística
- Extensão dos recursos para várias línguas (mapeamento dos recursos do OpenLogos – IT, FR, SP, GE)
- Desenvolvimento e melhoria do ParaMT
- Tradução automática

NECESSIDADE DE CRIAÇÃO DE REGRAS SINTÁCTICO-SEMÂNTICAS PARA A TRADUÇÃO

pregar Det N(missa,sermão) > preach = say N
pregar Det N(ideia,virtude,religião,verdade) > proclaim N > advocate N
pregar Det N(prego ,etc.) > hammer N(nail,etc.)
pregar N(susto) Prep(a) N > scare N
estabelecer Det N(negócio,empresa,loja,etc.) > open Det N
estabelecer N(regras,princípios) > lay down N
apresentar N(desculpa) > apologize
apresentar Det N(opinião,sugestão) > give Det N(opinion,suggestion)
apresentar Det N(moção,censura) > bring N forward
prestar Det N(serviço) > offer Det N(service)
prestar N(atenção) > pay N(attention)
perseguir N(objectivo,propósito,etc.) > follow N
perseguir N(pessoa) > chase N = hunt after/down N
pedir N(desculpa,perdão) > apologize = say sorry
pedir Det N(esmola) > beg
observar Det N(lua) > observe N
observar Det N(lei) > obey N(law)
provocar Det N(pessoa) > seduce N = make advances on N
provocar N(sarilho,confusão) > pick up N(trouble)
provocar N(tempestade) > cause N(storm)
representar Det N(papel) > play Det N(role)
representar Det N(contributo,etc.) > represent N

CONTEÚDO DA SESSÃO

❖ Onde estamos?

- ✓ O que é o **Port4NooJ**?
- ✓ Para que serve o **Port4NooJ**?

❖ Para onde caminhamos?

- ✓ Discussão sobre a melhoria e expansão do **Port4NooJ**
- ❑ Proposta de integração do **Port4NooJ** noutros recursos e de colaboração em projectos futuros que integrem o **Port4NooJ**

COLABORAÇÕES FUTURAS



CONTEÚDO DA SESSÃO

❖ Onde estamos?

- ✓ O que é o **Port4NooJ**?
- ✓ Para que serve o **Port4NooJ**?

❖ Para onde caminhamos?

- ✓ Discussão sobre a melhoria e expansão do **Port4NooJ**
- ✓ Proposta de integração do **Port4NooJ** noutros recursos e de colaboração em projectos futuros que integrem o **Port4NooJ**

PUBLICAÇÕES RELACIONADAS COM O **PORT4NOOJ**

- Belinda Maia & Anabela Barreiro "**Uma experiência de recolha de exemplos classificados de tradução automática de inglês para português**". In Diana Santos (ed.), Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa. Lisboa, Portugal:IST Press, 2007, pp. 205-216.
- Luís Sarmiento, Anabela Barreiro, Belinda Maia & Diana Santos "**Avaliação de Tradução Automática: alguns conceitos e reflexões**". In Diana Santos (ed.), Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa. Lisboa, Portugal:IST Press, 2007, pp. 181-190.
- Anabela Barreiro. "**ParaMT: a Paraphraser for Machine Translation**". In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.), Computational Processing of the Portuguese Language, 8th International Conference, Proceedings (PROPOR 2008) Vol. 5190, (Aveiro, Portugal, 8-10 de Setembro de 2008), Springer Verlag, pp. 202-211.
- Anabela Barreiro. "**Port4Nooj: Portuguese Linguistic Module and Bilingual Resources for Machine Translation**". In Proceedings of the 2007 International Nooj Conference (Barcelona, Spain, June 7-9, 2007), Cambridge Scholars Publishing.
- Anabela Barreiro & Elisabete Ranchhod. "**Machine Translation Challenges for Portuguese**". Linguisticæ Investigationes 28.1 (2005), pp. 3-18. (Machine Translation, Controlled Languages and Specialised Languages). Amsterdam/Philadelphia: John Benjamins Publishing Company. ISSN: 0378-4169.
- Anabela Barreiro "**Novas Ferramentas e Recursos Linguísticos para a Tradução Automática: Por ocasião d'O Fim do Início de uma Nova Era no Processamento da Língua Portuguesa**". In Luís Costa, Diana Santos & Nuno Cardoso (eds.), Perspectivas sobre a Linguatca / Actas do encontro Linguatca : 10 anos. Linguatca, 2008, pp. 13-23.

LINKS E LEITURA RECOMENDADA

Links

- <http://www.nooj4nlp.net/pages/nooj.html>
- <http://poloclup.linguatca.pt/Port4NooJ/> (neste sítio Web, recomenda-se vivamente a leitura do [Tutorial introdutório](#) (em português) e a [Descrição dos recursos](#) (em inglês))
 - Anabela Barreiro. "**Port4NooJ Linguistic Resources Overview**". 2008.
 - Anabela Barreiro. "**Tutorial introdutório do Port4NooJ**". 2008.

Outra leitura recomendada

- Silberztein, Max. 2003. [NooJ manual](#). Disponível no sítio do NooJ: <http://www.nooj4nlp.net>
- Silberztein, Max. 2005. [NooJ's Dictionaries](#). In the Proceedings of the 2nd Language and Technology Conference, Poznan University: 2005.
- Barreiro, Anabela. "Port4NooJ: an open source, ontology-driven Portuguese linguistic system with applications in machine translation". Oral presentation at the 2008 International NooJ Conference (NooJ'08) (Budapeste, Hungary, 8-10 June, 2008). [Slides](#)

Mais bibliografia (livros, actas, características específicas, tutoriais) em:

- <http://www.nooj4nlp.net/pages/references.html>

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela **Fundação para a Ciência e a Tecnologia**, através da bolsa de doutoramento com a referência SFRH/BD/14076/2003, co-financiada pelo POSI e pela **Fundação para a Computação Científica Nacional - Linguateca**, através de um contrato de trabalho de 1/2 meses para ajudar a tornar os recursos do **Port4NooJ** e o **ReEscreve** públicos. O contrato foi actualmente renovado por mais 3 meses para alargar e melhorar os recursos linguísticos aqui descritos.



PORT4NooJ

ATÉ ONDE SE PODE IR?

Anabela Barreiro*

Cristina Mota

LINGUATECA/FCCN

*CLUP